The current state and TRL assessment of unattended and left-behind object detection technology

Prepared by:
M. Lalonde, M. Derenne, L. Gagnon, D. O. Gorodnichy
Canada Border Services Agency
79 Bentley Avenue
Ottawa ON Canada
K1A 0L8

Project # PSTP 03-401BIOM

Scientific Authority:
Pierre Meunier
DRDC Centre for Security Science
613-992-0753

The scientific or technical validity of this Contract Report is entirely the responsibility of the Contractor and the contents do not necessarily have the approval or endorsement of the Department of National Defence of Canada.

Defence Research and Development Canada
Contract Report
DRDC-RDDC-2014-C292
September 2014

IMPORTANT INFORMATIVE STATEMENTS

PROVE-IT (FRiV) Pilot and Research on Operational Video-based Evaluation of Infrastructure and Technology: Face Recognition in Video project, PSTP 03-401BIOM, was supported by the Canadian Safety and Security Program (CSSP) which is led by Defence Research and Development Canada's Centre for Security Science, in partnership with Public Safety Canada. Led by Canada Border Services Agency partners included: Royal Canadian Mounted Police, Defence Research Development Canada, Canadian Air Transport Security Authority, Transport Canada, Privy Council Office; US Federal Bureau of Investigation, National Institute of Standards and Technology, UK Home Office; University of Ottawa, Université Québec (ÉTS).

The CSSP is a federally-funded program to strengthen Canada's ability to anticipate, prevent/mitigate, prepare for, respond to, and recover from natural disasters, serious accidents, crime and terrorism through the convergence of science and technology with policy, operations and intelligence.

© Her Majesty the Queen in Right of Canada, as represented by the Minister of National Defence, 2014

© Sa Majesté la Reine (en droit du Canada), telle que représentée par le ministre de la Défense nationale, 2014

Canada Border Services Agency
Agence des services frontaliers du Canada

Science and Engineering Directorate

Border Technology Division

Division Report: 2014-13 (TR)

July 2014

The Current State and TRL Assessment of Unattended and Left-Behind Object Detection Technology

M. Lalonde,
M. Derenne,
L. Gagnon,
D. O. Gorodnichy

PROTECTION • SERVICE • INTEGRITY

Canada

This page left intentionally blank

Abstract

This report presents a survey of the state of the art in automatic Unattended/Left-behind Objects Detection (ULOD) in various premises (metro stations, train stations and airports), followed by the technology readiness level (TRL) assessment based thereon. The survey overviews recent academic advances in this area and also focuses on current commercial offering in the form of a product evaluation. The evaluation is based on the methodology established in previous technical challenges that were put in place during international conferences.

Keywords: video-surveillance, video analytics, abandoned object, object left behind, object removal, technology readiness, performance evaluation, data-sets.

Community of Practice: Border and Transportation Security

Canada Safety and Security (CSSP) investment priorities:

1. Capability area: P1.6. Border and critical infrastructure perimeter screening technologies/ protocols for rapidly detecting and identifying threats.

1. Specific Objectives: O1. Enhance efficient and comprehensive screening of people and cargo (identify threats as early as possible) so as to improve the free flow of legitimate goods and travellers across borders, and to align/coordinate security systems for goods, cargo and baggage;

2. Cross-Cutting Objectives CO1. Engage in rapid assessment, transition and deployment of innovative technologies for public safety and security practitioners to achieve specific objectives;

3. Threats/Hazards F. Major trans-border criminal activity e.g. smuggling people/material

Acknowledgements

This work is done within the project PSTP-03-402BTS \PROVE-IT(VA)" funded by the Defence Research and Development Canada (DRDC) Centre for Security Science (CSS) Public Security Technical Program (PSTP) and in-kind contributions from CRIM by the following contributors:

- Dr. Dmitry Gorodnichy, Science & Engineering Directorate, Canadian Border Services Agency
- Dr. Marc Derenne, Dr. Langis Gagnon, Dr. Marc Lalonde, Computer Research Institute of Montreal (CRIM)

The feedback from project partners: RCMP, TC, CATSA, DRDC, UK HomeOffice, FBI is gratefully acknowledged.

Disclaimer

In no way do the results presented in this paper imply recommendation or endorsement by the Canada Border Services Agency, nor do they imply that the products and equipment identified are necessarily the best available for the purpose. The information presented in this report contains only the information available in public domain.

Release Notes

Context: This document is part of the set of reports produced for the PROVE-IT(VA) project. All PROVE-IT(VA) project reports are listed below.

1. Dmitry Gorodnichy, Jean-Philippe Bergeron, David Bissessar, Ehren Choy, Jacque Sciandra, "Video Analytics technology: the foundations, market analysis and demonstrations", Border Technology Division, Division Report 2014-36 (TR).

2. Dmitry O. Gorodnichy, Diego Macrini, Robert Laganiere, "Video analytics evaluation: survey of datasets, performance metrics and approaches", Border Technology Division, Division Report 2014-28 (TR).

3. D. Macrini, V. Khoshaein, G. Moradian, C. Whitten, D.O. Gorodnichy, R. Laganiere, "The Current State and TRL Assessment of People Tracking Technology for Video Surveillance applications", Border Technology Division, Division Report 2014-14 (TR).

4. M. Lalonde, M. Derenne, L. Gagnon, D. Gorodnichy, "The Current State and TRL Assessment of Unattended and Left-Behind Object Detection Technology", Border Technology Division, Division Report 2014-13 (TR).

Jointly with the PROVE-IT(FRiV) project (PSTP-03-401BIOM):

5. D. Bissessar, E. Choy, D. Gorodnichy, T. Mungham, ``Face Recognition and Event Detection in Video: An Overview of PROVE-IT Projects (BIOM401 and BTS402)", Border Technology Division, Division Report 2013-04 (TR).

6. D. Gorodnichy, E. Granger J.-P.Bergeron, D.Bissessar, E.Choy, T. Mungham, R. Laganiere, S. Matwin, E. Neves C. Pagano, M. De la Torre, P. Radtke, "PROVE-IT(FRiV): framework and results". Border Technology Division, Division Report 2013-10. Proceedings of NIST International Biometrics Performance Conference (IBPC 2014), Gaithersburg, MD, April 1-4, 2014. Online at http://www.nist.gov/itl/iad/ig/ibpc2014.cfm

The PROVE-IT(VA) project took place from August 2011 till March 2013. This document was drafted and discussed with project partners in March 2013 at the *Video Technology for National Security* (VT4NS) forum. The final version of it was produced in July 2014.

Appendices: This report is accompanied by appendix which includes the presentation related to this report at the VT4NS 2013 forum.

Contact: Correspondence regarding this report should be directed to DMITRY dot GORODNICHY at CBSA dot GC dot CA.

Table of Contents

1. Introduction

A left-behind (or abandoned) object is defined as “a non-living static entity that is part of the foreground and has been present in the scene for an amount of time greater than a predetermined threshold” [4]. One common way of separating foreground pixels from background pixels is by means of background modeling techniques, which have been around since the late 1990’s. The challenge is first to locate these foreground pixels reliably and ‘monitor’ them for some time; then they must be assigned to the class “abandoned object” and possibly linked to their owner, provided that they are true objects, not e.g. immobile bystanders.

Prior to being abandoned, an object carried by an owner is 1) “put” on the floor (or other surfaces), 2) “unattended”, i.e. the owner leaves the object on the floor and walks away at a distance greater than a predetermined spatial threshold, and 3) “abandoned” (i.e. left-behind) when the owner does not get back to the object after a predetermined time threshold.

This report surveys the state of the art of automatic Unattended/Left-behind Objects Detection (ULOD) in various premises (metro stations, train stations and airports). The survey presents recent academic advances in this area and also focuses on current commercial offering in the form of a product evaluation. The evaluation is based on the methodology established in previous technical challenges that were put in place during international conferences. However, it is not exhaustive, as it relies on trial versions available at the time of writing this report.

The paper is organized as follows. Section II gives a brief overview regarding the state-of-the-art knowledge on ULOD, including literature review, main challenges/conferences, past programs/initiatives and available public datasets. Section III gives information on the available commercial products and describes the methodology used to test some of them. Test results are presented in Section IV. The discussion on the TRL assessment of the technology based on the obtained results concludes the paper.

2. **Academic advances**

This section describes the academic contributions pertaining to the ULOD problem. Literature review of the past three years is presented. The datasets and evaluation practices are described.

2. 1 Literature review

The following techniques or algorithms are usually found in systems that have the capability to perform abandoned object detection:

- Background subtraction: simple approaches rely on a single background model, either to extract and track foreground blobs; others create a series of foreground masks and ‘image counters’ that track blobs’ time of life. One of the common approaches is based on computing two background

models [1] - one for short-term detection (updated every frame) and the other for long-term detection (updated every n frames). Regions in the image where a change is detected in the computed long-term foreground mask but not in the short-term mask become good candidates for an abandoned object.

- Other approaches do not use background modeling at all ([2],[3]).
- A non-trivial difficulty that faces detection systems is the ability to find and maintain a relationship between the abandoned object and its owner, so that a piece of luggage is not flagged as abandoned when its owner is nearby. People tracking, possibly initialized by a person detector, helps verify the existence of this relationship.
- Static bystanders who occasionally move their head or limbs may inadvertently break the system. Many techniques of varying complexity have been proposed: from living/non-living detectors that make sure that candidate objects have stable contours ([4]), to more elaborate schemes involving a person detector ([5]).
- One technique shared by the most sophisticated systems found in the scientific literature is the finite-state automaton, which is used to track the state of a foreground object throughout its life and even beyond, when it eventually fuses with the background ([6],[7]).
- Researchers have also examined the contours of blobs associated with potentially abandoned objects in order to label the objects as being abandoned or removed (stolen). Active contours or segmentation/region growing are common tools that have been used in this context.

2.2 Main challenges/conferences

Three main conferences have been welcoming contributions in abandoned object detection in the recent years: PETS, AVSS and TRECVID (although this one is rather focusing on "object put").

The PETS conference (Performance Evaluation of Tracking and Surveillance) exists since 2000, and as the name implies, the objective is to encourage the evaluation of visual tracking and surveillance algorithms as low-level tasks. In 2006, a higher level task was proposed in the form of a challenge [8]: researchers were invited to work on abandoned object detection using a standard dataset of varying difficulty. Seven papers were presented at the conference in that context. In 2007, the theme was multi-sensor event recognition in crowded public areas [9], and again the provided dataset focused on i) loitering, ii) attended luggage removal ("theft"), and iii) left-luggage scenarios, of increasing complexity. Five papers reported progress in these areas.

The IEEE conference series on advanced video and signal based surveillance (AVSS) has been held episodically since 1998. The broad focus includes topics such as image processing, video processing, signal processing, audio processing, pattern recognition, and computer vision. Interestingly, the industry takes much room in this conference, as sponsors but also as scientific contributors and demonstrators/exhibitors. In 2007 [10], the conference hosted a challenge on abandoned item detection using a dataset from i-LIDS [11].

TRECVID is a well-known conference in the academic world [12]. Since 2003, this NIST-sponsored event has stimulated research by proposing tough challenges in information extraction from images and

videos: detection of objects in videos (cars, mountains, US flags, fire, etc.), detection of copied material, video shot detection, extraction of camera motion, etc. In 2008, a new task on video surveillance was launched, where participants were asked to design systems/algorithms capable of extracting security-related events in i-LIDS videos. Among the targeted events were people running or pointing, people getting together or splitting up, etc., as well as an event called 'object put' that occurs when someone puts a bag or a suitcase on the floor. Such an event is semantically close to the task of abandoned item detection that we are interested in.

2.3 Programs and initiatives

Many research programs and/or initiatives have been established to support research in video surveillance and some of them directly or indirectly target abandoned object detection. Let us mention SUBITO [13], Vanaheim [14], Samurai [15], ISCAPS [16], i-LIDS [11] and STIDP [17].

2.4 Available public datasets

The main datasets used in the scientific literature are those that were made available to participants in international challenges.

The PETS 2006 dataset contains seven scenarios of varying complexity ranging from 1 (easy) to 5 (very difficult), four cameras per scenario. Videos have been acquired by fairly good, consumer-type, cameras at PAL resolution. Video length is about 120 seconds. The PETS 2006 dataset should be considered somewhat easy because recent papers about abandoned object detection claim 100% detection (or close) on this dataset.

The PETS 2007 dataset contains nine scenarios, including four 'theft' (luggage taken away from the owner) and two 'unattended' (owner walks away), with varying difficulty. Again, four cameras monitor the same location. Video acquisition was done with the same equipment as for PETS2006.

The AVSS 2007 dataset contains three instances of the abandoned baggage scenario: 'easy', 'medium' and 'hard'. Movie clips are drawn from the i-LIDS dataset. Video length is short (e.g. 3min 30sec for the 'easy' clip).

The TRECVID contains 144 hours of videos acquired from five cameras at Gatwick Airport and containing hundreds of instances of the 'ObjectPut' event. Each camera monitors a different location. The dataset is real footage as opposed to the scripted scenarios of PETS.

The CANDELA [20] is a small dataset acquired during development of a specific subtask of the large CANDELA project (Content Analysis and Network DELivery Architectures; 2003-2005; 15 participants; budget >15M €). The subtask was about abandoned object detection.

Finally, the CAVIAR dataset [21] is a large dataset with tens of video files associated to some scenarios. Five instances of the 'Leaving bags behind' scenario are publicly available for download. Similarly to CANDELA, the level of activity in the scene is low.

A subset of representative situations from the PETS and AVSS datasets was used in the current evaluation.

3. Product evaluation

3.1 Commercial products

A significant number of commercial vendors in the automatic video surveillance market propose products that have the ability to perform abandoned object detection. A non-exhaustive list includes AgentVI (Israel/USA), DVTel/ioimage (USA), Intellio (Hungary), VCA Technology (UK), IntuVision (USA), i2V (India), Bikal (USA), Nice Systems (Israel), Bellsent (China), IOmniscient (Autralia), 2020 Imaging (UK), Shyam Network (India), Total Imaging Solution (USA), Bosch Security (Germany), BTCO (Saudi Arabia), Fibridge (China), AIT (Austria), Ipsotec (UK) and Evitech (France).

For the evaluation, we were able to download four[1] products from the Internet, identified by the letters A, B, C and D. These products were shipped with a trial license that typically lasted 15 or 30 days. Although time and budget limitations prevented us from purchasing and deploying a large number of solutions, the four demos are good quality products and are representative of the current market offering in video surveillance. Here are some notes and observations about these products:

- Systems A, B and C accept a file (e.g. AVI) as a video source; for system D, a program called webcamXP was used to act as a virtual IP camera (http protocol, MJPEG video format).
- Although being a demo version, product B was shipped with an hour of free technical support; we took advantage of this opportunity to ask for assistance in tuning the system for the AVSS sequence.
- Product A was the only one exporting the list of alarm events in an XML file; the others did not provide this functionality and thus the protocol included a manual stage (transcription of the results) that slowed down the evaluation.

3.2 Evaluation methodology

In order to conduct a credible and non-subjective evaluation, we reused the methodology adopted for the AVSS 2007 challenge [18].

- It considers an abandoned baggage as a non-moving object that was brought inside the detection area by a person who then left the area without it for at least *n* seconds (*n* fairly high, e.g. 45-60 seconds).
- Alarm events are compared to ground truth data according to Figure 1.
- True positive alarms as well as false negatives and false positives allow the computation of recall and precision, and ultimately the F1 score computed as follows:

$$\text{F1 score}=(k+1)\cdot\text{Recall} \cdot \text{Precision}/(\text{Recall}+k\cdot\text{Precision}),$$

where Recall=$a/(a+c)$, Precision=$a/(a+b)$,

a is the true positive alarms, b the false positive alarms and c the false negative alarms.

[1] A fifth product with good reputation was considered but 1) no evaluation version was available, probably because of the complexity of the system and 2) maintaining contact with the vendor was difficult, and time ran out before we could get a quotation from them.

Finally, it should be underlined that the evaluation is based on a temporal alignment of detected events without consideration for spatial information, which means that a situation depicted in Figure 2 will be regarded as a valid detection of an abandoned object even though the detection is clearly erroneous. At first sight, one might be surprised with this loose definition of a valid event, but the end result is a notification to the officer in charge of the surveillance system who will carry out the alarm validation task, regardless of the spatial accuracy of the detection. In the future, new generations of the systems with stronger detection capabilities may require evaluation criteria that take spatial accuracy into account, so that the system that assist the officer in accurately locating the abandoned object get a better score.

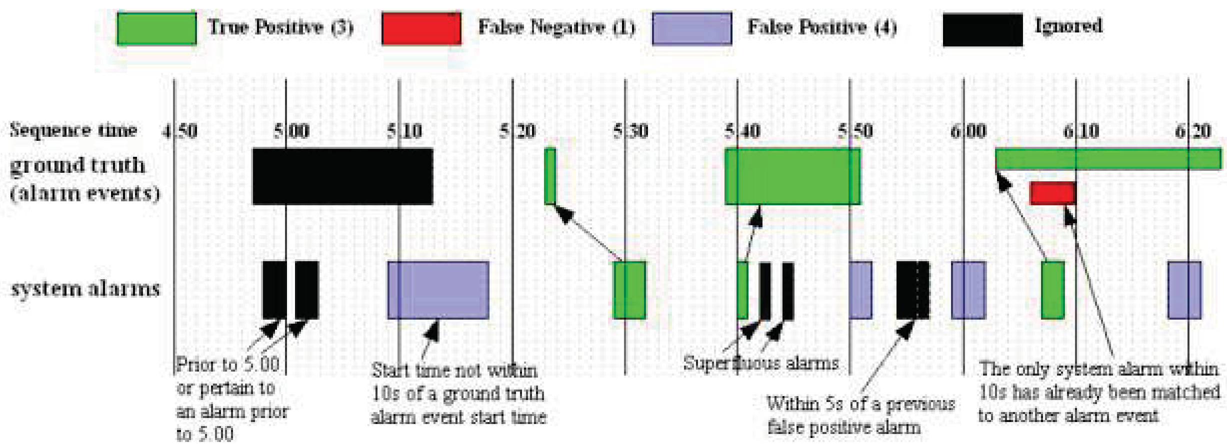


Figure 1. Event alignment w.r.t. ground truth (drawn from the AVSS challenge guide)

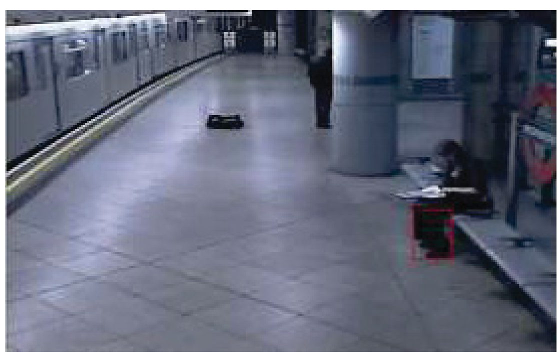


Figure 2. Accidental detection of abandoned luggage

3.3 Dataset and ground truth

The dataset used for the evaluation is a mix of video sequences used for the AVSS 2007, PETS 2006 and PETS 2007 challenges. The videos have been selected because 1) they represent de facto standard data in the research community, and 2) they are available on the Internet, so the product evaluation can be reproduced easily. The names of the sequences as well as their length of time are listed in Table I.

The difficulty rating for AVSS videos is related to the location of the luggage in the scene: a piece of luggage close to the camera (at the bottom of the frame in Figure 2) appears bigger with less people walking or standing in front of it, so this situation is considered an easy case; on the other hand, an item that is much farther from the camera is only a few tens of pixels wide in the video frame and there is strong likelihood that it will get occluded if crowd density increases, so this situation is rated as 'hard'.

PETS 2007 videos appear in gray boxes in Table 1 because they are not suitable for scoring. The reason is related to the definition of 'abandoned object' given for the corresponding challenge, which states that an object is abandoned (unattended) when its owner is at a certain minimal distance from it. PETS 2007 videos feature people walking around their luggage and actually never leaving the scene, so these data are not really compatible with the 'luggage left behind' scenario described in the previous Section. We will still use them for qualitative assessment of system performance in detecting static objects.

TABLE I. VIDEO SEQUENCES USED FOR TESTING

From AVSS	From PETS2006	From PETS 2007
AVSS2007 Easy (3m38s)	PETS2006 S2-T3-C cam1 (1m25s; difficulty 3/5)	PETS2007_s07_1stview (1m40s; difficulty 2/5)
AVSS2007 Medium (3m13s)	PETS2006 S2-T3-C cam2	PETS2007_s07_2ndview
AVSS2007 Hard (3m32s)	PETS2006 S2-T3-C cam3	PETS2007_s07_3rdview
AVSS2007 Eval (21m45s)	PETS2006 S2-T3-C cam4	PETS2007_s07_4thview
	PETS2006 S7-T6-B cam1 (1m53s; difficulty 5/5)	PETS2007_s08_1stview (1m40s; difficulty 4/5)
	PETS2006 S7-T6-B cam2	PETS2007_s08_ 2ndview
	PETS2006 S7-T6-B cam3	PETS2007_s08_3rdview
	PETS2006 S7-T6-B cam4	PETS2007_s08_4thview

Extensive ground truthing has been done at CRIM over the selected sequences. Each 'abandoned luggage' event has been annotated as follows:

- ObjectPut (time at which the object is left/put down).
- PersonMovedAway (time at which the owner starts moving away from the object).
- PersonLeftScene (time at which the owner becomes invisible, presumably unable to look after the object).

- PersonIsBack (time at which the owner is in the camera view again).
- ObjectPickup (time at which the owner removes the object from the scene).
- Alarm Duration.

Of interest, of course, is the PeopleLeftScene time, at which point an alarm should be triggered.

Some modifications have been made to the AVSS videos: the first black frames with introductory text have been replaced with a fixed image of the empty scene so as to allow some systems to jumpstart background adaptation efficiently.

For a similar reason, videos from the CANDELA project are not part of the evaluation dataset due to their length of time (between 12 and 49 seconds) which may be too short for adequate system adaptation.

4. Results

This section describes the results obtained following evaluation of the four commercial products A, B, C and D. For each product, the AVSS Easy and AVSS Medium sequences were used to manually tune the various system parameters which then remained constant for the rest of the evaluation. It can be argued that additional parameter tuning would have been necessary for each type of sequence (AVSS, PETS2006, PETS2007) because of differences in camera positioning, scene appearance, crowd density, lighting, etc.; indeed, results are expected to be the most significant for AVSS_Hard and AVSS_Eval, whereas results for PETS can be viewed as indicators of system flexibility and ease of use/configuration/deployment.

Subsections A and B contain tables of results for two subsets of the whole dataset, namely the videos from AVSS2007 and PETS2006. Abbreviations for some column labels are as follows: TP=true positives, FP=false alarms, FN=false negatives, Prec=precision. Note that the entries in the TP column that appear as “0*” represent good spatial detections of an abandoned item but with a bad timing when compared to the ground truth.

4.1 Results for AVSS2007

TABLE II. **RESULTS FOR SYSTEM A (AVSS DATA)**

Video	Detections			Prec.	Reca ll	F1
	TP	FP	FN			
AVSS_Easy	1	0	0	1.0	1.0	1.0
AVSS_Medium	1	0	0	1.0	1.0	1.0
AVSS_Hard	1	0	0	1.0	1.0	1.0
AVSS_Eval	1	11	5	0.083	0.167	0.16

TABLE III. RESULTS FOR SYSTEM B (AVSS DATA)

Video	Detections			Prec.	Recall	F1
	TP	FP	FN			
AVSS_Easy	0*	1	1	0	0	-
AVSS_Medium	0	0	1	0	0	-
AVSS_Hard	0	1	1	0	0	-
AVSS_Eval	1	21	5	0.045	0.167	0.16

TABLE IV. RESULTS FOR SYSTEM C (AVSS DATA)

Video	Detections			Prec.	Recall	F1
	TP	FP	FN			
AVSS_Easy	0	1	1	0	0	-
AVSS_Medium	0	1	1	0	0	-
AVSS_Hard	0	1	1	0	0	-
AVSS_Eval	1	8	5	0.11	0.167	0.16

TABLE V. RESULTS FOR SYSTEM D (AVSS DATA)

Video	Detections			Prec.	Recall	F1
	TP	FP	FN			
AVSS_Easy	0	1	1	0	0	-
AVSS_Medium	0	2	1	0	0	-
AVSS_Hard	0	3	1	0	0	-
AVSS_Eval	3	16	3	0.16	0.5	0.47

4.2 Results for PETS 2006

TABLE VI. RESULTS FOR SYSTEM A (PETS 2006 DATA)

Video	Detections			Prec.	Recall	F1
	TP	FP	FN			
S2-T3-C cam1	0*	1	1	0	0	-
S2-T3-C cam2	0	0	1	-	0	-
S2-T3-C cam3	0*	1	1	0	0	-
S2-T3-C cam4	0*	0	1	-	0	-
S7-T6-B cam1	0*	1	1	0	0	-
S7-T6-B cam2	1	0	0	1.0	1.0	1.0
S7-T6-B cam3	0*	1	1	0	0	-
S7-T6-B cam4	0*	1	1	0	0	-

TABLE VII. RESULTS FOR SYSTEM B (PETS 2006 DATA)

Video	Detections			Prec.	Recall	F1
	TP	FP	FN			
S2-T3-C cam1	0	2	1	0	0	-
S2-T3-C cam2	0	3	1	0	0	-
S2-T3-C cam3	0	3	1	0	0	-
S2-T3-C cam4	0	1	1	0	0	-
S7-T6-B cam1	0	6	1	0	0	-
S7-T6-B cam2	0	1	1	0	0	-
S7-T6-B cam3	0	1	1	0	0	-
S7-T6-B cam4	0	2	1	0	0	-

TABLE VIII. **RESULTS FOR SYSTEM C (PETS 2006 DATA)**

Video	Detections			Prec.	Recall	F1
	TP	FP	FN			
S2-T3-C cam1	0	0	1	-	0	-
S2-T3-C cam2	0	0	1	-	0	-
S2-T3-C cam3	0	0	1	-	0	-
S2-T3-C cam4	0	0	1	-	0	-
S7-T6-B cam1	0	0	1	-	0	-
S7-T6-B cam2	0	1	1	0	0	-
S7-T6-B cam3	0	0	1	-	0	-
S7-T6-B cam4	0	0	1	-	0	-

TABLE IX. **RESULTS FOR SYSTEM D (PETS 2006 DATA)**

Video	Detections			Prec.	Recall	F1
	TP	FP	FN			
S2-T3-C cam1	0*	1	1	0	0	-
S2-T3-C cam2	0*	1	1	0	0	-
S2-T3-C cam3	0	0	1	-	0	-
S2-T3-C cam4	0	0	1	-	0	-
S7-T6-B cam1	0	0	1	-	0	-
S7-T6-B cam2	1	1	0	0.5	1.0	0.97
S7-T6-B cam3	0	0	1	-	0	-
S7-T6-B cam4	0	0	1	-	0	-

4.3 Results for PETS 2007

As mentioned earlier, results for PETS 2007 video sequences are qualitative. The figures below show screen shots containing detections found by the products under evaluation. For systems A and B, images from the third view seem to be processed more efficiently but it is no surprise that the detections appear to be more accurate (at least spatially):

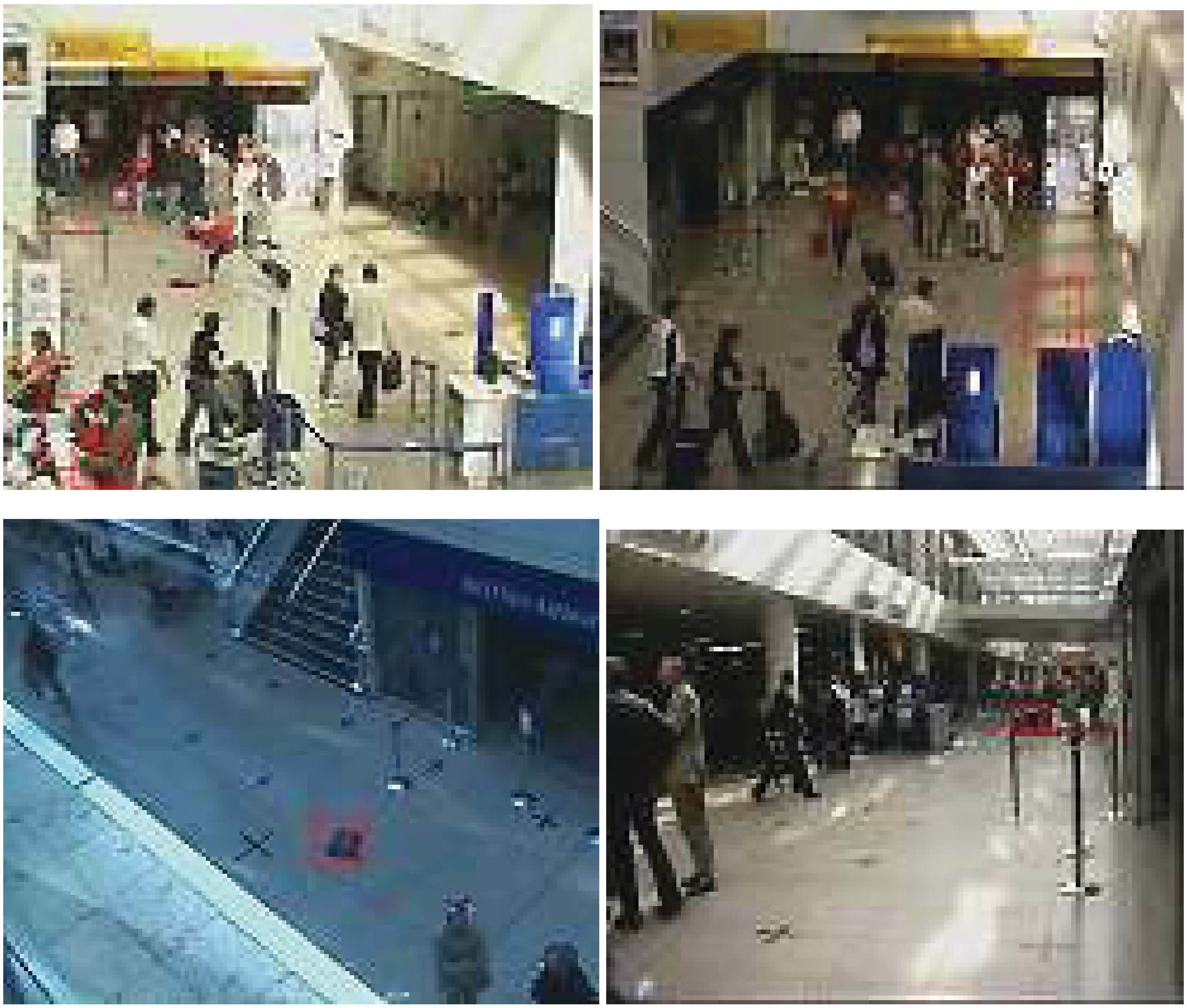


Figure 3. Four views from Sequence S7 with System A (views are not necessarily time-aligned).

Figure 4. Sequence S7 with system C (only one detection on 3rd view; no detection with S8)

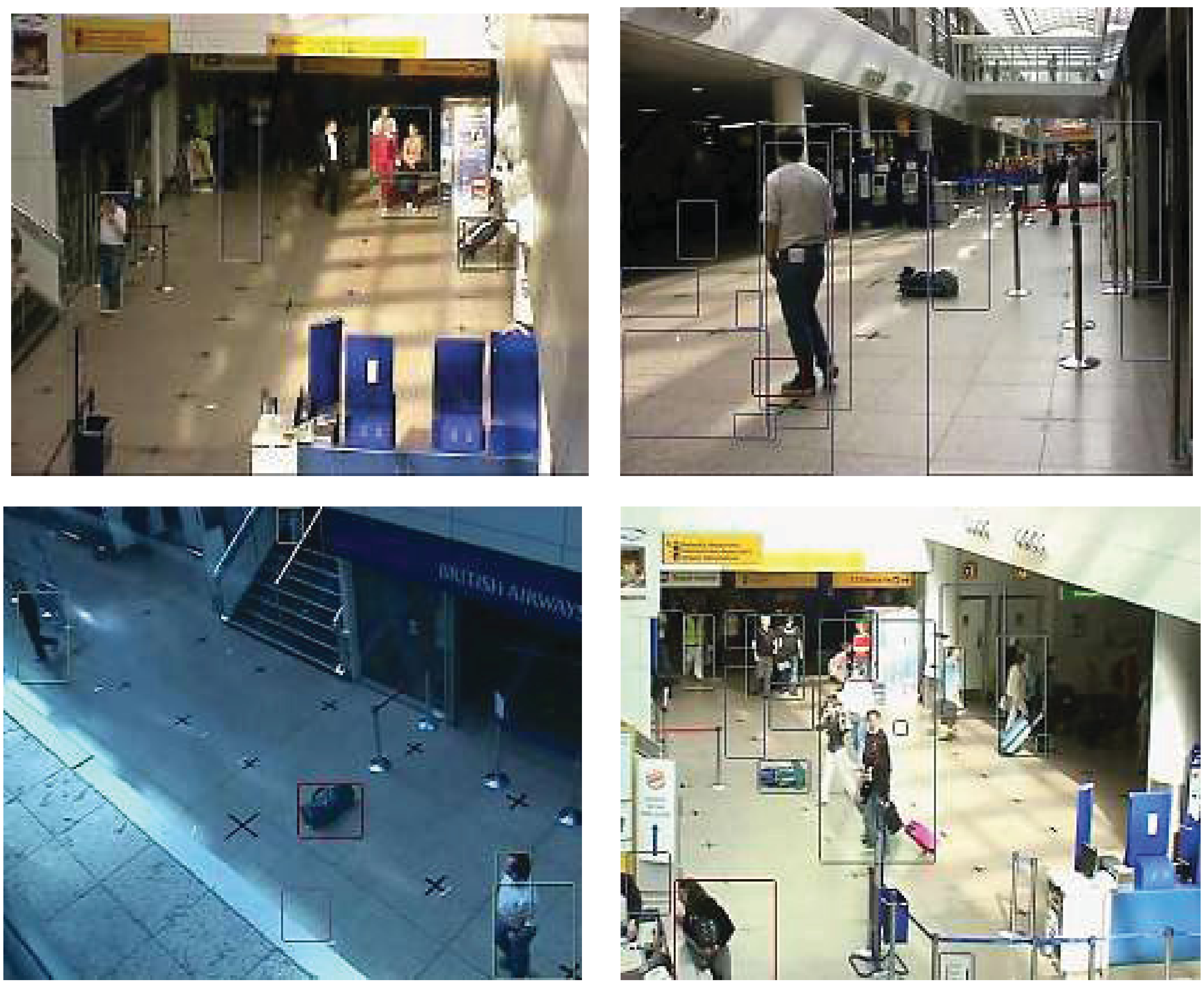


Figure 5. Four views from Sequence S8 with System B; no detection on second view (views are not necessarily time-aligned).

4.4 Issues

One issue that quickly arises is the mismatch between the requirement stated in the description of the methodology, namely the ability to detect objects that have been abandoned by the owner, and what current commercial systems offer as a feature, which is closer to static object detection. Concretely, an alarm should be raised as the bag owner leaves the area but the systems under evaluation have simpler rules that raise an alarm *n* seconds after an object has been abandoned, even though the owner might be around, looking after the object from a reasonable distance. Of course, such a mismatch will have a negative impact on the F1 scores collected during the experiment even though event alignment rules depicted in Figure 1 are designed to be relatively insensitive to small differences between event alarm start time and the ground truth. But the mismatch also reflects the difficulty of designing a system that should handle such a high-level semantic event: detecting that an item is abandoned when its owner leaves the scene means that the system should be aware of the relationship between the owner and the object and should be able to monitor it, a capability that implies person tracking and possibly person re-identification in case of a crowded scene. Yet few reliable solutions exist in the research community for these problems. Two excerpts from the final report of the SUBITO project [19] tends to underline the same finding:

- "The experimental results achieved demonstrated that the inclusion of reasoning about the intentions of individuals within a scene and the interactions between these individuals leads to greatly improved performance over the state of the art in abandoned baggage detection"
- "A competitive assessment was also carried out comparing SUBITO functionality to similar product offerings in the current market place. It was found that the SUBITO system capabilities exceeds those of most deployed systems (products) and uniquely exploits the "concept of ownership" principle that is fundamental to effective threat management and resolution."

Figure 6. One example of object flagged as abandoned, although owner is nearby.

Figure 7. Another example of item wrongly flagged as abandoned.

Apart from the functionality mismatch discussed in the previous paragraph, two limitations of the evaluation procedure can (partially) explain the high error rates recorded.

1. Limited time and expertise in parameter tuning for each specific product is an obvious limitation. Some systems have a large set of parameters and better results might have been obtained if an expert had spent time adjusting the various thresholds. Thanks to the support people behind product B who graciously did some tuning for a few video sequences, it has been possible to qualitatively assess the sensitivity of the parameters that influence the behavior of this product. A related factor is camera placement: some systems perform better when using overhead cameras, but this recommendation cannot be followed with imposed datasets.
2. Another, more subtle, limitation is related to the length of the video sequences being used. During product testing we noticed that some systems may consume a large number of video frames during initialization (e.g. up to 45s for system D). The influence on system behavior and performance cannot be categorized as negligible because some correct detections have been recorded with systems running in 'replay mode' only (same video, but never ending) despite the fact that no detection had occurred during the first pass.

5. TRL Assessment

5.1 Introduction to TRL Assessment

The results from an empirical evaluation of a technology (or an application), such as those measured in terms of False / True Negatives and Positives and their derivatives - Precision and Recall, while being very informative from academic point of view and providing the basis for comparing one product to another, cannot be easily used by an operational agency that needs to know whether or not a technology is ready for deployment. This is why operational communities prefer evaluating a technology / application in terms of the Technology Readiness Levels (TRL) [22,23], which ranges from Level 1 to Level 9:

- Level 1 (Basic principles observed and reported),
- Level 2 (Technology concept and/or application formulated)
- Level 3 (Analytical and experimental critical function and/or characteristic proof of concept.
- Level 4 (Component validation in laboratory environment),
- Level 5 (Laboratory-scale similar system or component validated in relevant environment),
- Level 6 (Pilot-scale similar prototypical system or component validated in relevant environment),
- Level 7 (Full-scale prototypical system demonstrated in relevant environment),
- Level 8 (Actual system completed and qualified through test and demonstration),
- Level 9 (Actual system successfully operated in the field over the full range of expected conditions).

TRL assessment is adopted by many agencies as a risk management tool. It provides a common scale of science and technology exit criteria and allows one to estimate the cost/investment required for deploying a system.

Conducting TRL assessment requires forming a team of unbiased and independent subject-matter experts, who set the protocol and the schedule for evaluating the technology (or application), and who, having obtained the sufficient amount of technical evidence, collectively decide on the TRL of the technology / application in question. The detailed description of all TR levels and the process for conducting a TRL assessment, including the supporting information required for each level, is available at [23, Section 2.5].

5.2 Application of TRL Assessment

Table X presents the assessment on the TRL for Unattended / Left-Behind Object Detection technology, based on the observations and findings obtained through the course of this study and the discussion of those results and findings with our government and academic stakeholders and project partners.

Assessment is done for complete technology components, as well as for technology sub-components, for different types of environmental and scenario factors and settings, using three levels:

- "-": substantially not suitable for pilot or deployment
- "?": maybe suitable for further investigation or pilot
- "+" suitable for a live pilot or mock-up simulation testing.

The decision to use the three-grade metric instead of the original nine-grade TRL scale is due to the intent of the study to serve as a starting reference point for a more detailed analysis on the technologies in question, rather than to provide an ultimate verdict on the technology readiness, which may not be possible within the limited timeframe and resources available for the study.

TABLE X. **ASSESSMENT ON THE TRL FOR UNATTENDED / LEFT-BEHIND OBJECT DETECTION (SEE BOTTOM FOR LEGEND)**

Detection Technology	**Type 1**	**Type 2a**	**Type 2b**	**Type 3**	**Type 4**
Carried Object	**-**	**-**	**-**	**-**	**-**
Dropping Object (Object Put)	**?** **+V**	**?** **+STV**	**-**	**-**	**-**
Static Object for more than *n* sec.	**+** **+**	**+** **+V**	**?** **+STV**	**-**	**-**
Unattended Object	**?** **+STV**	**?** **+STV**	**-**	**-**	**-**
Abandoned Object	**?** **+STV**	**?** **+STV**	**-**	**-**	**-**
Object Removal (Object Picking)	**?** **+STV**	**?** **+STV**	**-**	**-**	**-**
Person-baggage Association	**?** **+STV**	**-**	**-**	**-**	**-**
Owner Change	**-**	**-**	**-**	**-**	**-**

Type 1: Primary Inspection Lane (PIL) kiosk, Passport Control
Type 2a: controlled chokepoint (one person at a time following the same direction)
Type 2b: uncontrolled chokepoint (many persons at a time following the same direction)
Type 3: indoor uncontrolled (airport, metro stations, etc.)
Type 4: free flow outdoors

For each surveillance setup of increasing complexity (Type 1…Type 4), TRL is given for each of the following conditions:
“S” (Object Size): small (<1/32th of the image width) vs. large
“T” (Traffic): little (< 20 moving objects per 1 min per 1/32 of image width) vs. dense
“V” (Viewing conditions, e.g. occluded often): good vs. challenging

For example:
+ → works for all conditions
+T → works only at little traffic
+S → works only for large objects
+V → works only for good viewing condition
+ST→ works only for large objects and in little traffic only

The detection technologies are sorted in their usual chronological actions: carried object, dropping object, stationary object, unattended object, abandoned object, followed possibly by object removal (retaking or picking) and owner change. For this last action, person-baggage association is essential and is in fact critical in the context of all the above abandoned object actions. Only when this task is achievable can a real ULOD system be built. In the meantime, current ULOD systems are mainly capable only of detecting objects being static for more than *n* seconds rather than detecting an abandoned or left-behind object.

In general, the main limitations in building a reliable ULOD system are the following: intense scene activity (creating occlusions), lack of person-baggage association, small object and illumination change. Commercial systems appear to perform reasonably well when these limitations are not present. However, based on the literature survey results and the tests conducted, it strongly believed that these systems will not work satisfactorily in more challenging environments without generating many false alarms, and are therefore substantially not suitable for pilot / further investigation (marked as “-“ in Table X).

6. Discussion

We have presented a survey on automatic Unattended/Left-behind Objects Detection (ULOD) in various premises (metro stations, train stations and airports). We have covered recent academic advances in this area and current commercial offering in the form of a product evaluation and TRL assessment. The evaluation was based on the methodology established in previous technical challenges that were put in place during international conferences. However, it is not exhaustive as it relies on trial versions available at the time of writing this report.

It is important to note that TRL is only one out of several technology maturity metrics used by operational communities. Others include, as adopted from [24]:

1. **Producibility or Manufacturing Readiness**, which relates to the readiness of industry to produce the technology,
2. **User Readiness or Practice Based Technology Maturity,** which emphases the readiness of end-users to receive and operate the technology,
3. **Program Readiness**, which relates to the business needs to receive the technology and the ability to develop business requirements and procedures for deploying the technology,
4. **Research & Development (R&D) Readiness,** which relates to the R&D capacity required to customize and tune the technology for the field requirements

It is therefore important to consider all technology maturity metrics when making a decision about the deployment of a technology, especially if the technology is new and does not have a proven success record history.

In conclusion, one should be careful when interpreting the published results because, even though the reported results may be justified and sound (including datasets and metrics), the implementation of the same technology in a different context, e.g. as part of a complex, multi-module commercial system or in a different surveillance settings, may lead to the results that are more worse than ones reported.

References

[1] F. Porikli, Y. Ivanov, and T. Haga, "Robust Abandoned Object Detection Using Dual Foregrounds," EURASIP Journal on Advances in Signal Processing, vol. 2008, Article ID 197875, 11 pages, 2008

[2] S. Kwak, G. Bae and H. Byun, "Abandoned luggage detection using a finite state automaton in surveillance video", Opt. Eng. 49, 027007 (Mar 02, 2010).

[3] B. K. Mitra, W. Hassan, P. Birch, A. Gardezi, R. C. D. Young, C. R. Chatwin, "A two-stage approach to detect abandoned baggage in public places, "Proc SPIE Visual Information Processing XIX, 6 April 2010, Orlando, Florida, USA. Volume 7701 of SPIE Proceedings, pages 770107, SPIE, 2010.

[4] D. Bhatnagar, "Dropped object detection in crowded scenes," MIT M.S. Thesis, Dept. of Electrical Engineering and Computer Science, 2009.

[5] H. Liao, J. Chang, and L. Chen, "A Localized Approach to Abandoned Luggage Detection with Foreground-Mask Sampling", Proc. AVSS 2008, pp.132-139.

[6] Q. Fan and S. Pankanti, "Modeling of Temporarily Static Objects for Robust Abandoned Object Detection in Urban Surveillance," Proc. 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), 2011

[7] R. H. Evangelio and T. Sikora, "Static Object Detection Based on a Dual Background Model and a Finite-State Machine," EURASIP Journal on Image and Video Processing, vol. 2011, Article ID 858502, 11 pages, 2011

[8] http://www.cvg.rdg.ac.uk/PETS2006/index.html

[9] http://pets2007.net/

[10] http://www.eecs.qmul.ac.uk/~andrea/avss2007_d.html

[11] http://www.homeoffice.gov.uk/science-research/hosdb/i-lids

[12] http://trecvid.nist.gov/

[13] http://www.subito-project.eu/Home.htm

[14] http://www.vanaheim-project.eu/

[15] http://www.samurai-eu.org/Home.htm

[16] http://www.iscaps.reading.ac.uk/home.htm

[17] http://www.dhs.gov/xlibrary/assets/privacy/privacy-pia-st-stidp.pdf

[18] http://www.eecs.qmul.ac.uk/~andrea/dwnld/avss2007/AVSS_2007_i-LIDS_challenge_briefing.pdf

[19] http://www.subito-project.eu/PDF/SUBITO_Progress_Report_3_Summary.pdf

[20] http://www.hitech-projects.com/euprojects/cantata/datasets_cantata/dataset.html

[21] http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/

[22] Mankins, John C. (6 April 1995). "Technology Readiness Levels: A White Paper". NASA, Office of Space Access and Technology, Advanced Concepts Office. http://www.hq.nasa.gov/ office/codeq/trl/trl.pdf.

[23] "Technology Readiness Assessment (TRA) Guidance". United States Department of Defense. April 2011. http://www.acq.osd.mil/ chieftechnologist/publications/docs/TRA2011.pdf

[24] Dmitry Gorodnichy. "Video Analytics: Technology Maturity, Deployment Challenges, and Roadmap", Proceedings of the Third Interdepartmental Workshop on Video Technologies for National Security - VT4NS'10. Available at the DRDC SharePoint portal, also as Appendix D of Division Report 2014-36 (TR) Analytics technology: the foundations, market analysis and demonstrations".

Appendix A: "Estimating the TRL of an unattended / left-behind baggage detection system" (Presentation from VT4NS'13)

Estimating the TRL of an unattended / left-behind baggage detection system

Marc Lalonde
Langis Gagnon
CRIM

Ottawa
March 27, 2013

Partenaire financier :
Développement économique, Innovation et Exportation Québec

Outline

- Introduction
- Tests with commercial products
  - Product characteristics, evaluation methodology, datasets
  - Results: AVSS2007, PETS2006, PETS2007
- A word on TRECVID-SED
- Discussion / estimating the TRL
- Conclusion

Introduction

- Left-behind object: “non-living static entity that is part of the foreground and has been present in the scene for an amount of time greater than a predetermined threshold”
- Abandoned baggage: “non-moving object that was brought inside the detection area by a person who then left the area without it.”

Introduction (cont.)

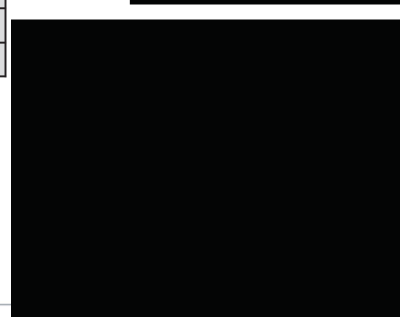
- Difficult task:
  - Need for foreground/background separation.
  - Occlusions, static bystanders.
  - Ability to find and maintain a relationship between the abandoned object and its owner.
  - Key factors: object size, level of activity in scene.
- Objective: assess readiness level (TRL) of this technology through evaluation of commercial products.

Products vs. methodology

- Four products selected based on availability of trial version:
  - Systems labeled A, B, C, and D.
- Their task: detect abandoned objects.
- Event comparison to ground truth according to AVSS 2007 challenge.
- Evaluation based on precision-recall, F_1 score.

Dataset

AVSS2007 Easy (3m38s)	PETS2006 S2-T3-C cam1 (1m25s; difficulty 3/5)	PETS2007_s07_1stview (1m40s; difficulty 2/5)
AVSS2007 Medium (3m13s)	PETS2006 S2-T3-C cam2	PETS2007_s07_2ndview
AVSS2007 Hard (3m32s)	PETS2006 S2-T3-C cam3	PETS2007_s07_3rdview
AVSS2007 Eval (21m45s)	PETS2006 S2-T3-C cam4	PETS2007_s07_4thview
	PETS2006 S7-T6-B cam1 (1m53s; difficulty 5/5)	PETS2007_s08_1stview (1m40s; difficulty 4/5)
	PETS2006 S7-T6-B cam2	PETS2007_s08_ 2ndview
	PETS2006 S7-T6-B cam3	PETS2007_s08_3rdview
	PETS2006 S7-T6-B cam4	PETS2007_s08_4thview



CRIM

Results: AVSS 2007

'*' means good detection with bad timing

A

Video	Detections			Prec.	Recall	F1
	TP	FP	FN			
Easy	1	0	0	1.0	1.0	1.0
Medium	1	0	0	1.0	1.0	1.0
Hard	1	0	0	1.0	1.0	1.0
Eval	1	11	5	0.083	0.167	0.16

B

Video	Detections			Prec.	Recall	F1
	TP	FP	FN			
AVSS_Easy	0*	1	1	0	0	-
AVSS_Medium	0	0	1	0	0	-
AVSS_Hard	0	1	1	0	0	-
AVSS_Eval	1	21	5	0.045	0.167	0.16

C

Video	Detections			Prec.	Recall	F1
	TP	FP	FN			
AVSS_Easy	0*	1	1	0	0	-
AVSS_Medium	0	1	1	0	0	-
AVSS_Hard	0*	1	1	0	0	-
AVSS_Eval	1	8	5	0.11	0.167	0.16

D

Video	Detections			Prec.	Recall	F1
	TP	FP	FN			
AVSS_Easy	0*	1	1	0	0	-
AVSS_Medium	0	2	1	0	0	-
AVSS_Hard	0	3	1	0	0	-
AVSS_Eval	3	16	3	0.16	0.5	0.47

CRIM

Results : PETS2006

A

Video	Detections			Prec.	Recall	F1
	TP	FP	FN			
S2-T3-C cam1	0*	1	1	0	0	-
S2-T3-C cam2	0	0	1	-	0	-
S2-T3-C cam3	0*	1	1	0	0	-
S2-T3-C cam4	0*	0	1	-	0	-
S7-T6-B cam1	0*	1	1	0	0	-
S7-T6-B cam2	1	0	0	1.0	1.0	1.0
S7-T6-B cam3	0*	1	1	0	0	-
S7-T6-B cam4	0*	1	1	0	0	-

B

Video	Detections			Prec.	Recall	F1
	TP	FP	FN			
S2-T3-C cam1	0	2	1	0	0	-
S2-T3-C cam2	0	3	1	0	0	-
S2-T3-C cam3	0*	3	1	0	0	-
S2-T3-C cam4	0*	1	1	0	0	-
S7-T6-B cam1	0	6	1	0	0	-
S7-T6-B cam2	0	1	1	0	0	-
S7-T6-B cam3	0	1	1	0	0	-
S7-T6-B cam4	0*	2	1	0	0	-

C

Video	Detections			Prec.	Recall	F1
	TP	FP	FN			
S2-T3-C cam1	0	0	1	-	0	-
S2-T3-C cam2	0	0	1	-	0	-
S2-T3-C cam3	0	0	1	-	0	-
S2-T3-C cam4	0	0	1	-	0	-
S7-T6-B cam1	0	0	1	-	0	-
S7-T6-B cam2	0	1	1	0	0	-
S7-T6-B cam3	0	0	1	-	0	-
S7-T6-B cam4	0	0	1	-	0	-

D

Video	Detections			Prec.	Recall	F1
	TP	FP	FN			
S2-T3-C cam1	0*	1	1	0	0	-
S2-T3-C cam2	0*	1	1	0	0	-
S2-T3-C cam3	0	0	1	-	0	-
S2-T3-C cam4	0	0	1	-	0	-
S7-T6-B cam1	0	0	1	-	0	-
S7-T6-B cam2	1	1	0	0.5	1.0	0.97
S7-T6-B cam3	0	0	1	-	0	-
S7-T6-B cam4	0	0	1	-	0	-

Results : PETS2007

s07_1stview

B+8 f/a

s07_2ndview

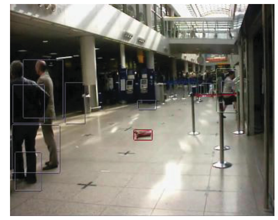
No detection

s07_3rdview

A+2 f/a

B+0 f/a

s07_4thview

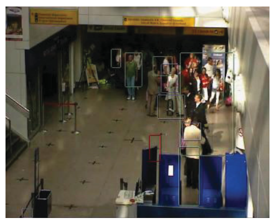


B+2 f/a

s08_1stview

A+3 f/a

s08_2ndview



No detection

s08_3rdview

B+1 f/a

s08_4thview

A+1 f/a; B: 0 det., 9 f/a

Results (cont.)

- 2-hr sequence with very low scene activity, no abandoned luggage (TRECVID : LGW_20071101_E1_CAM4):

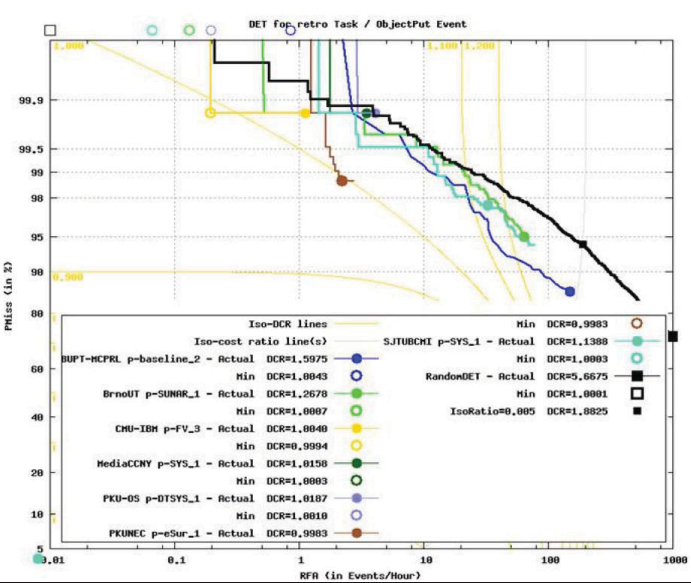
System	# f/a
A	0
B	1
C	1
D	2-3

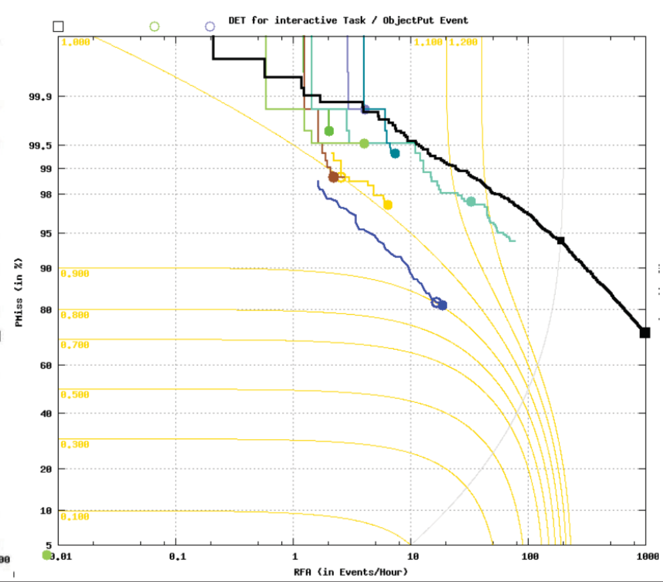
Results (cont.)

- Notes of caution:
  - Non-optimal parameter tuning
  - Length of test sequences may be a problem

A word on TRECVID-SED

- Task called ‘ObjectPut’ has close relationship with detection of abandoned objects.
- Yet after 5 rounds of competition, results are not too good:





CRIM

Discussion

- Main issue: mismatch between what is expected and what products can deliver.
  - Expected: detection of abandoned objects
  - Delivered: detection of objects being static for more than n seconds.

CRIM

Discussion (cont.)

- Key finding: commercial products have no concept « luggage ↔ owner »

- Re-identification?

www.crim.ca
ISO 9001:2008

CRIM

Estimating the TRL

"S" (Object Size): small (<1/32th of the image width) vs. large
"T" (Traffic): little (< 20 moving objects per 1 min per 1/32 of image width) vs. dense
"V" (Viewing conditions, e.g. occluded often): good vs. challenging

VA detection technology	Type 1 fixed light person lane	Type 2 fixed light small crowd	Type 3 fixed light large crowd	Type 4 variable light small crowd	Type 5 variable light large crowd
Carried Object	-	-	-	-	-
Dropping Object	? [4]	?, +STV [4]	-	? [4]	-
Static Object for more than n sec.	✔ [7?]	✔, +V [6-7?]	?, +STV [4]	? [4]	-
Unattended Object	?, +STV [4]	?, +STV [4]	-	-	-
Abandoned Object	?, +STV [4]	-	-	-	-
Object left behind	? [4]	?, +STV [4]	-	-	-
Person-baggage Association	? [4]	-	-	-	-
Owner Change	-	-	-	-	-

CRIM

Estimating the TRL (cont.)

- Availability of commercial products = hint that TRL could be high.
- In fact, these products can do a decent job at detecting static objects in simple situations.
- Bottom line:
  - TRL < 4 for most scenarios.
  - TRL = 4 in simple cases (people flow).
  - Static object detection has higher probability of success.

CRIM

Conclusion

- Four products tested according to AVSS2007 methodology
- As expected: no concept of ownership, false alarms (higher rate with increasing scene activity)
- In most cases, TRL<4.